

ABSTRACT OF THE DISCLOSURE

A natural language translation method and system translating medical reports created in natural language into structured data frames that can be utilized in computer databases for decision support, billing, research, and other purposes. Structured data entry is elicited from a patient in order to identify an appropriate disease signature corresponding to his or her condition and symptoms. In turn, the disease signature identifies the appropriate lexical domain with which to analyze the natural language report. The translation method and system use statistical analysis based on empirical data that particular combinations of words have interdepended previously within a modeled context and how frequently individual words interdepend generally and with what kinds of words. For each sentence in the report, the words in the medical report are looked up in the lexical domain individually and in combination with all other words coexisting in the same sentence. The word combinations are parsed to determine the likelihood the words interdepend in the report. For those words determined to interdepend, a semantic interpreter defines the semantic relationship between the words. A frame generator compiles the word relationships into records having fields recognized as pertinent by the disease signature and that can be searched and sorted by computers on those fields.